Confirmatory Factor Analysis of the BCSSE Scales

Justin Paulsen, ABD

James Cole, PhD

January 2019

Indiana University
Center for Postsecondary Research
1900 East 10th Street, Suite 419
Bloomington, Indiana 47406-7512
Phone: 812-856-5824
Email: bcsse@indiana.edu

TABLE OF CONTENTS

INTRODUCTION 1

METHOD

Data Source and Characteristics 1

BCSSE Scales 2

RESULTS

Descriptive Analysis 3

Internal Consistency Reliability 4

Confirmatory Factor Analysis 9

SUMMARY 10

REFERENCES 12

APPENDICES

Appendix A: Items Included in the BCSSE Scales 13

Introduction

The Beginning College Survey of Student Engagement (BCSSE, pronounced "bessie") measures entering first-year students' high school academic and co-curricular experiences as well as their expectations for participating in educationally purposeful activities during the first year of college. Since BCSSE's launch in 2007, nearly 900,000 entering first-year students enrolled at 506 institutions have participated in the survey.

BCSSE was updated in 2013 to align with the updated version of NSSE. The new version maintained BCSSE's focus on gathering information from entering first-year students regarding their high school experiences and their expectations for engagement during their first year in college. It also included new items to increase alignment with NSSE, improve clarity and applicability of survey language, and refine existing measures. The revision also included new items and nine scales (aka, BCSSE scales). One of the goals during the revision process was to assure that these scales have strong psychometric properties. Much of the development work was completed during the NSSE update process and included extensive student cognitive interviews, literature reviews, expert consultations, pilot testing, statistical analysis of pilot data, and interviews with administrators responsible for use of the BCSSE data on their campuses. BCSSE was updated again in 2019 to include transfer and delayed-entry students (older students with no prior college experience) for the online survey mode. This report only evaluates the psychometric properties of BCSSE scales using data from first-year students. A future psychometric report will include data from transfer and delayed-entry students. However, pilot data collected from transfer and delayed-entry students in 2018 did not indicate any deviation in scale internal consistency compared to first-year students.

Two statistical techniques were used to reevaluate the psychometric properties of the BCSSE Scales (See Cole & Dong, 2014). First, internal consistency reliability was investigated using Cronbach's alpha. The second technique used a split-sample approach to identify possible factor structures and the second to test the identified structure.

Method

Data Source and Characteristics

The data for this study are drawn from the 2018 administration of the BCSSE survey of entering first-year students completing the survey enrolled at 118 U.S. institutions and one Canadian University. Approximately 36% of these first year students completed the paper version of the survey, with the remaining 64% completing the survey online. Of the 118 US institutions, 36% are baccalaureate level, 43% masters, and 17% doctoral. Approximately 65% of the participating institutions are private. Approximately 40% of these students were first-generation college students. Fifty-nine percent of the respondents identified as a woman, 40% as a man, 1% as another gender or prefer not to respond. Regarding the students' race/ethnicity, <1% identified as American Indian/Alaskan Native, 8% as Asian, 11% as Black/African American, 9% as Hispanic, 60% as White, 11% as multiracial, and 1% as other.

Given that BCSSE is locally administered on each campus, precise response rates are not calculable for BCSSE. However most institutions attempt to administer the BCSSE to their entire entering first-year class. For those institutions, completion rates ranged from 13% to 100%, with an average completion rate of 70%.

BCSSE Scales

A total of 42 items were categorized into nine BCSSE Scales (Table 1). The total number of items in each scale ranged from 3 to 7 (see Appendix A for complete list of items). Scale values ranged from 0 to 60. These scales were developed using exploratory and confirmatory factor analysis from pilot data from NSSE (2011-2012) and during the initial pilot testing of BCSSE (2004-2006).

Table 1. BCSSE Scales

High School Quantitative Reasoning (HS_QR)	Expected Academic Perseverance (EXP_PER)
High School Learning Strategies (HS_LS)	Expected Academic Difficulty (EXP_DIF)
Expected Collaborative Learning (EXP_CL)	Perceived Academic Preparation (PER_PREP)
Expected Student-Faculty Interaction (EXP_SFI)	Importance of Campus Environment (IMP_CAMP)
Expected Discussions with Diverse Others (EXP_DD)	

RESULTS

Descriptive Analysis

Five of the BCSSE Scales (HS_QR, HS_LS, EXP_CL, EXP_SFI, EXP_Dif) have means within 10 points of the mid-point 30 (Table 2). The other four have means that are greater than 10 points beyond the scale mid-point (EXP_DD, EXP_PER, EXP_PREP, IMP_CAMP), with three of them (EXP_PER, EXP_PREP, and EXP_CAMP) having the smallest standard deviations. The scale EXP_DD has the highest mean (44.8) and second largest standard deviation. Overall, these scales have adequate distributional properties with the possible exception of EXP_DD, which is skewed toward the upper end of the range due to the larger standard deviation.

The confidence intervals (CIs) are estimates of the "true" population value (Barde & Barde, 2012). Wider CIs indicate less precision, while narrower ones indicate greater precision. The 95% confidence intervals for the BCSSE scales range in size from .15 to .22 (Table 2). These results comport with analysis of BCSSE data from the previous 3 years where ranges in confidence intervals for each scale never exceeded .22. During this same four year period (2015 to 2018), the average absolute scale mean difference ranged between .19 (2017 to 2018) and .83 (2015 to 2017). These small year-to-year differences in CIs and scale means assures that we can be 95% confident that the true mean for each scale is very close to sample means.

Table 2. Descriptive statistics for each BCSSE Scale.

						95% Conf Int	
	N	M	SD	Skewness	Kurtosis	Lower	Upper
HS_QR	67646	30.5	14.60	.118	-.487	30.38	30.60
HS_LS	67559	38.2	13.00	-.092	-.564	38.11	38.30
EXP_CL	66719	37.9	11.57	.097	-.470	37.82	38.00
EXP_SFI	66493	33.6	12.97	.329	-.440	33.47	33.66
EXP_DD	66219	44.8	13.63	-.540	-.339	44.65	44.86
EXP_PER	65961	42.6	9.67	-.429	.309	42.51	42.66
EXP_DIF	65749	28.6	10.22	.107	.177	28.57	28.72
PER_PREP	65378	43.4	9.73	-.356	.119	43.35	43.50
IMP_CAMP	65435	45.6	9.94	-.606	.140	45.55	45.70

Internal Consistency Reliability

Internal consistency reliability for this study was measured by a) Cronbach's alpha for each measure, b) Cronbach's alpha for a measure if a single item is removed, c) correlations between an item and the remaining items in the measure (called corrected item-scale correlations), d) the average inter-item correlation, the range of inter-item correlations, and e) the individual inter-item correlations of the scale. All correlations are Pearson's r correlations. The criteria used are summarized in Table 3.

Cronbach's alpha measures the internal consistency of a group of items by measuring the homogeneity of the group of items. "It is an indication of how well the different items complement each other in their measurement of different aspects of the same variable or quality" (Litwin, 2003, p. 22). Cronbach's alpha ranges in value between zero and one. Values closer to one indicate a higher internal consistency; values closer to zero indicate a lower internal consistency. McMillan and Schumacher (2001) suggest that groups of items with an alpha below .70 should be used with caution. Cronbach's alpha if an item is deleted provides a measure at the item level as to whether that item contributes to the scale's internal consistency. If the scale's alpha measure increases after removing the item, the item is not a meaningful contributor to the scale's internal consistency.

The internal consistency of a scale can also be examined with item-to-scale correlations and intercorrelations of items within a scale (DeVellis, 2003). If a group of items measures a single latent construct, we would assume that each item alone correlates with the scale overall and that items within such a scale are positively correlated. According to Clark and Watson (1995), average inter-item correlations should fall somewhere between .15 and .50 as anything below .15 would suggest the proposed construct is too broad while anything above .50 would indicate redundancy of items on the scale. Additionally, a corrected item-scale correlation examines the correlation between the individual item and the scale score excluding the given item. This measure indicates whether someone scoring high on an individual item also scores high on the overall scale. Positive correlations are expected and should be at least 0.30 and higher (de Champlain, 2009; Varma, 2006).

Table 3. Internal Consistency Criteria for This Study

Reliability Statistics	Criteria for a Good Scale
Cronbach's Alpha	Greater than or equal to .70
Range of inter-item correlations	between .15 and .85
Average inter-item correlation	Between .15 and .50
Range of Cronbach's alpha's if item deleted	Deleting any item would decrease the alpha
Range of corrected item-scale correlations	Greater than or equal .30

Cronbach's alphas for the BCSSE scales and average inter-item correlations can be found in Table 4. The results in Table 4 suggest a high degree of internal consistency for most of the nine BCSSE scales. Cronbach's alphas range between .66 and .91 with all but two of our scales (HS_LS and EXP_DIF) above our criteria of .70. The inter-item correlations are all between .15 and .85. The average inter-item correlations rose well above .50 only in the case of Expected Discussions with Diverse Others, while High School Quantitative Reasoning and Expected Student-Faculty Interactions are slightly above 0.50. This indicates that those scales, particularly Expected Discussions with Diverse Others, contain items that are strongly intercorrelated meaning that students don't strongly differentiate between the items on the scale. This is reasonable as Expected Discussions with Diverse Others and High School Quantitative Reasoning are the narrowest constructs in the survey. Expected Student-Faculty Interactions may be particularly narrow as incoming students may have only a vague conception of the nature of student-faculty relationships at the college level and may not be able to discern among the different kinds of interactions there can be. No average inter-item correlation fell below .15 indicating that none of the scales represent overly broad constructs.

Table 4. Scale Cronbach's Alphas

BCSSE scales	Cronbach's α	Inter-Item Correlation	Average Inter-Item Correlation
High School Quantitative Reasoning	.78	.50-.62	.55
High School Learning Strategies	.68	.32-.57	.41
Expected Collaborative Learning	.79	.39-.66	.49
Expected Student-Faculty Interaction	.85	.52-.63	.58
Expected Discussions with Diverse Others	.91	.63-.82	.72
Expected Academic Difficulty	.66	.22-.48	.33
Expected Academic Perseverance	.81	.26-.55	.43
Perceived Academic Preparation	.86	.35-61	.47
Importance of Campus Environment	.85	.27-.80	.46

The range of each scale's overall Cronbach's alpha if individual items are removed and the range of corrected item-scale correlations can be found in Table 5. Individual item-scale analyses can be found in Table 6. With the exception of a couple of items (hLSreading and fyacadexp), all items meet the criteria for a good scale, i.e., if they were deleted the scale's Cronbach alpha would decrease. For these two items, the increase in reliability is marginal (+.01). None of the items fall below the threshold for corrected item-scale correlation, suggesting that the individual items correlate well with the overall scores on the given scale.

Overall, these nine BCSSE scales show moderate to high levels of internal consistency. Results suggest that High School Engagement in Learning Strategies and Expected Academic Difficulty are somewhat less internally consistent, which may be a result of shorter scales and lower average inter-item correlation. Expected Discussions with Diverse Others, and High School Quantitative Reasoning and Expected Student-Faculty Interactions to a lesser extent, are more narrowly focused scales. Researchers wanting the most internally consistent scales may want to consider removing the hLSreading or fyacadexp items from the High School Learning Strategies and Importance of Campus Environment scales respectively. Overall results suggest, however, that these nine BCSSE scales can be considered reliable measures of incoming students' expectations for and perceptions of their coming undergraduate student engagement.

Table 5. Scale Item-Scale Analyses

BCSSE Scales	Cronbach's α If Item Deleted	Corrected Item-Scale Correlation
High School Quantitative Reasoning	.67-.76	.57-.66
High School Learning Strategies	.48-.72	.38-.58
Expected Collaborative Learning	.70-.78	.52-.67
Expected Student-Faculty Interaction	.79-.81	.66-.71
Expected Discussions with Diverse Others	.87-.91	.73-.84
Expected Academic Difficulty	.54-.61	.41-.51
Expected Academic Perseverance	.76-.81	.47-.68
Perceived Academic Preparation	.82-.85	.55-.76
Importance of Campus Environment	.82-.86	.43-.74

Table 6. Individual Item Analyses

Scale	Items	Cronbach's α If Item Deleted	Corrected Item-Scale Correlation
High School Quantitative Reasoning	hqrconclud	.76	.57
	hqrproblm	.67	.66
	hqrevaluat	.68	.64
High School Learning Strategies	hLSreading	.72	.38
	hLSnotes	.52	.54
	hLSsummry	.48	.58
Expected Collaborative Learning	fyCLaskhlp	.75	.59
	fyCLxplain	.78	.52
	fyCLstudy	.70	.67
	fyCLprojct	.73	.62
Expected Student-Faculty Interaction	fySFcareer	.81	.66
	fySFothrwrk	.81	.67
	fySFprform	.79	.71
	fySFdiscuss	.81	.68
Expected Discussions with Diverse Others	fyDVrace	.89	.80
	fyDVeconomc	.87	.84
	fyDVreligion	.88	.83
	fyDVpolitical	.91	.73
Expected Academic Difficulty	clearnma	.60	.42
	cmantime	.61	.41
	cgethelp	.54	.51
	cintfac	.60	.41
Expected Academic Perseverance	cotherint	.80	.51
	cfindinfo	.78	.61
	ccourdis	.78	.60
	caskinst	.78	.61
	cfinish	.76	.68
	cstaypos	.81	.47
Perceived Academic Preparation	fySGwrite	.84	.63
	fySGspeak	.84	.64
	fySGthink	.82	.76
	fySGanalyze	.84	.61
	fySGothers	.84	.59
	cgncompt13	.85	.55
	cgninq	.84	.63
Importance of Campus Environment	fyacadexp	.86	.43
	fySEacad	.84	.60
	fySEdiv	.83	.66
	fySEnacad	.84	.58
	fySEsoc	.82	.74
	fySEact	.82	.71
	fySEserv	.83	.63

Confirmatory Factor Analysis

The confirmatory factor analysis was constructed based on the factor structure identified in previous research. In preparation for the confirmatory factor analyses, the BCSSE 2018 dataset was randomly divided in half. The first half of the sample was used to identify possible correlations among items and the second to test the identified structure. First, a confirmatory factor analysis was conducted using the AMOS 25.0 statistical software program based on the previously identified factor structure. We iteratively fit a model for each scale to identify correlations among scale items using one of the randomly drawn datasets. We included possible modifications by considering common meanings in survey items, magnitude of modification indices, standardized residuals, and improvements in model fit indices CFI, TLI, and RMSEA. After adopting certain modifications (see appendix A for finalized models[1]), we fit the given model with the remaining half of the dataset to ensure that selected modifications and resulting goodness of fit indicators are not sample specific. We report the results from this final fit.

We modeled CFAs for each of the individual scales as well as for all of the scales at the same time. For the overall model, we allowed all of the latent scale factors to correlate with each other. We assessed fit with criteria commonly used (Kenny, 2014) across CFA studies: RMSEA (root mean square error of approximation), TLI (Tucker Lewis Index), and CFI (comparative fit index). Typical cutoffs for each of these measures have been prescribed that suggest the relative goodness of fit. An RMSEA less than 0.01 indicates excellent fit, less than 0.05 indicates good fit, and less than 0.08 indicates mediocre fit; TLI and CFI greater than 0.95 suggests good fit, between 0.95 and 0.90 suggests marginal fit, and less than 0.90 suggests poor fit (Kenny, 2014). We report the 90th percentile confidence interval for RMSEA to provide greater detail. Scales with only three items are just identified, and so there are no measures of fit for those scales.

Based on these criteria, Table 7 suggests that all the models fit have marginal or good fit depending on the measure being considered. The RMSEA for the overall model suggests a good fit, while the TLI and CFI suggest a marginal fit. The RMSEA for EXP_PER and IMP_CAMP suggest a mediocre fit, while the TLI and CFI for those scales suggest good fit. The RMSEA for PER_PREP suggests poor fit, the TLI suggest marginal, while the CFI suggests good fit. These

[1] The overall model is not shown, but it consists of each of the individual scales as shown in the Appendix with the addition of correlations among all of the individual scales.

results suggest this scale might be the least coherent in terms of internal structure. The other scales all suggest excellent fit.

Table 7: Summary of Fit Indices from the Confirmatory Factor Analyses

BCSSE Scales	RMSEA 90th Percentile Confidence Interval	TLI	CFI	Total N
Overall Model	(0.045, 0.046)	0.901	0.911	58,163
HS_QR	N/A	N/A	N/A	65,980
HS_LS	N/A	N/A	N/A	65,830
EXP_CL	(0.000, 0.009)	1.000	1.000	64,419
EXP_SFI	(0.033, 0.052)	0.993	0.999	64,026
EXP_DD	(0.025, 0.043)	0.998	1.000	64,074
EXP_PER	(0.064, 0.071)	0.963	0.983	63,488
EXP_DIF	(0.006, 0.024)	0.998	1.000	63,532
PER_PREP	(0.089, 0.095)	0.937	0.967	62,903
IMP_CAMP	(0.073, 0.078)	0.960	0.979	62,812

Table 8 below presents the standardized regression weights from the overall model CFA. Standardized regression weights indicate the strengths of the factor loadings. The standardized regression weights suggest good strength of loading should they be greater than 0.40 (Kline, 2002). With the exception of one item loading (cintfac – EXP_DIF), all factor loadings showed strong loading on the given factor meaning that they are reliable indicators of the given factor.

Summary

Overall, the evidence suggests that researchers should feel confident treating the scale scores as latent constructs. Between the goodness of fit criteria and the individual item standardized regression weights, each of the scales have meaningful evidence suggesting that the items in the scale cohere to form a valid factor. The preponderance of evidence reported here should provide confidence to researchers, first-year program staff, and administrators who use BCSSE data.

Table 8: Overall Model Standardized Regression Weights

Scale	Item	Standardized Regression Weights
HS_QR	hqrconclud	.662
	hqrproblm	.782
	hqrevaluat	.772
HS_LS	hlsreading	.491
	hlsnotes	.702
	hlssummry	.763
EXP_CL	fyclaskhlp	.589
	fyclxplain	.535
	fyclstudy	.841
	fyclprojct	.771
EXP_SFI	fysfcareer	.775
	fysfothrwrk	.728
	fysfprform	.783
	fysfdiscuss	.794
EXP_DD	fydvrace	.881
	fydveconomc	.928
	fydvreligion	.829
	fydvpolitical	.719
EXP_PER	cstaypos	.499
	cfinish	.735
	caskinst	.705
	ccourdis	.688
	cfindinfo	.657
	cotherint	.541
EXP_DIF	clearnma	.661
	cmantime	.632
	cgethelp	.474
	cintfac	.342
IMP_CAMP	fyseserv	.677
	fyseact	.713
	fysesoc	.750
	fysenacad	.649
	fysediv	.732
	fysenacad	.633
	fysediv	.488
	fyseacad	.681
	fyacadexp	.488
PER_PREP	cgninq	.681
	cgncompt13	.548
	fysgothers	.663
	fysganalyze	.587
	fysgthink	.804
	fysgspeak	.699
	fysgwrite	.679

References

Barde, M. P., & Barde, P. J. (2012). What to use to express the variability of data: Standard deviation or standard error of mean? Perspectives in Clinical Research, 3, 113-116.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.

Cole, J. S., & Dong, Y. (2014). *Confirmatory factor analysis of the BCSSE Scales*. National Survey of Student Engagement, Indiana University.

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical education, 44*, 109-117.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

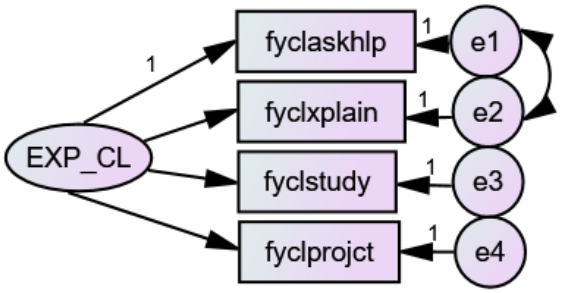
Litwin, M. S. (2003). *How to assess and interpret survey psychometrics* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

McMillan, J. H., & Schumacher, S. (2001). *Research in education: A conceptual introduction*. New York: Longman.

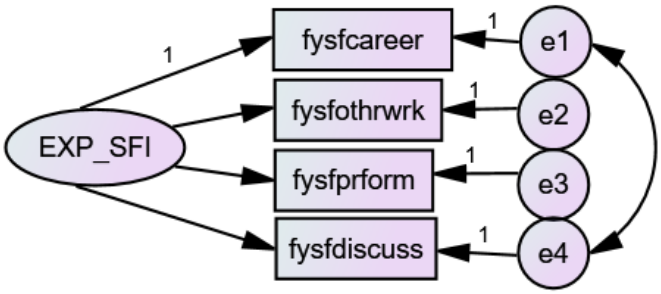
Varma, S. (2006). *Preliminary item statistics using point-biserial correlation and p-values*. Educational Data Systems Inc.: Morgan Hill CA. Retrieved from https://eddata.com/wp-content/uploads/2015/11/EDS_Point_Biserial.pdf

Appendix A: CFA Models

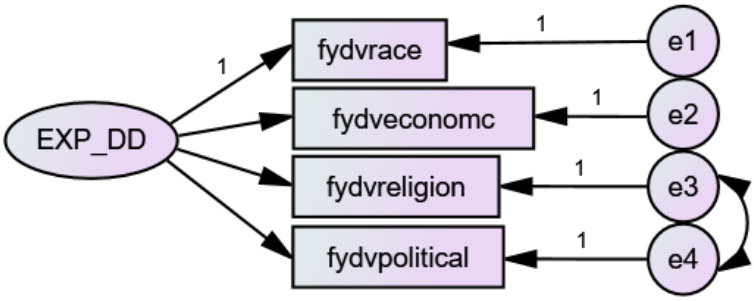
Expected Collaborative Learning



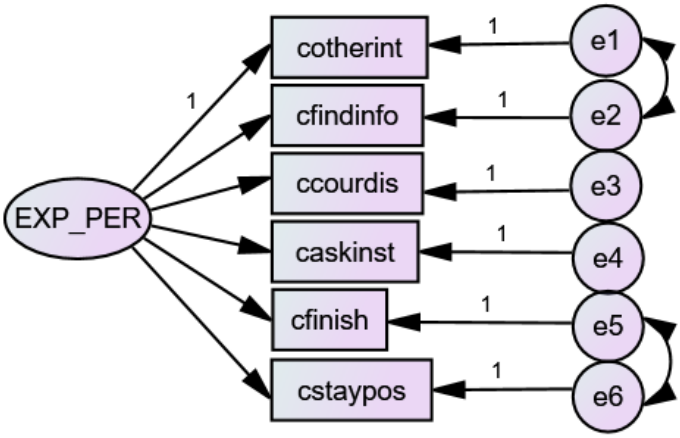
Expected Student-Faculty Interaction



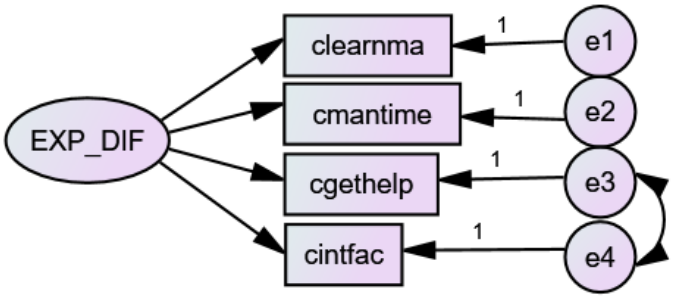
Expected Discussions with Diverse Others



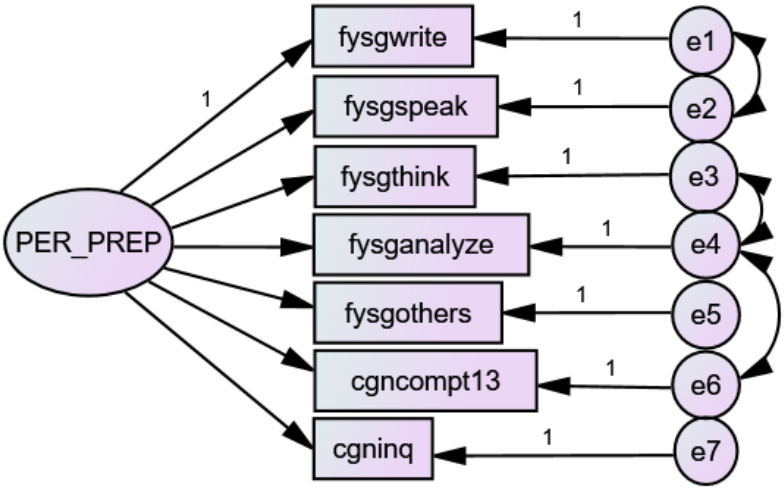
Expected Academic Perseverance



Expected Academic Difficulty



Perceived Academic Preparation



Importance of Campus Environment